Percent Similarity: 53.333 Percent Identity: 32.000

```
alignment_block:
 US-08-487-032A-764 x AA024217/rev   ..

 Align seg 1/1  to reverse of: AA024217  from: 1  to: 586

      84 AlaLeuAsnGlyPheGlyLeuAsnValGlyTyrLysLysPhePheGlnPh 100
         :::|||:::|||||||||||||                  :::   ||
     413 GGGCTTCACGGTTTTGGCCTT.....................CTTGGGTT 385

     100 eLysSerLeuAspMetThrSerLysTrpPheGlyPheArgValTyrGlyL 117
         |   :::||||||           ||||||:::|||        ||||
     384 TGGAGGCCTTGAC...........TGGTTTGACTTTGACAACCTTGGGC 347

     117 euPheAspTyrGlyHisAlaAspLeuGlyLysGlnValTyrAlaProAsn 133
         |||||:::|||             |||||||:::              :::
     346 TTTTTGGCTAT.....CTTGGGCGT...........GGCAGC 321

     134 LysIleGlnLeuAspMet.....................ValSerTr 142
         ::::::   |||::::::                     ::::::||
     320 CGGCTTCTTCTTGGCCTTCTTGACAGGGGTGGCTTTGGGTTTCTTGCTTG 271

     142 pGlyValGlySer***LeuLeuAla 150
         ||||   |||:::||||||||||:::
     270 GGGCTTTGGAGGCAGCCTTCTTGGG 246
```

seq_name: gb_est19:AI048339

```
seq_documentation_block:
LOCUS        AI048339      516 bp    mRNA            EST       08-JUL-1998
DEFINITION   uh59c02.r1 Soares mouse embryonic stem cell NMES Mus musculus cDNA
             clone 1749698 5' similar to gb:X03473_cds1 HISTONE H1' (HUMAN);
             gb:X13171 Mouse mRNA for H1 histone subtype H1(0) (MOUSE);, mRNA
             sequence.
ACCESSION    AI048339
NID          g3296626
KEYWORDS     EST.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
             Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 516)
  AUTHORS    Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
             Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
             Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
             Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
             Waterston,R.
  TITLE      The WashU-HHMI Mouse EST Project
  JOURNAL    Unpublished (1996)
COMMENT
             Contact: Marra M/Mouse EST Project
             WashU-HHMI Mouse EST Project
             Washington University School of MedicineP
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: mouseest@watson.wustl.edu
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             MGI:961510
             Seq primer: -28m13 rev2 ET from Amersham
             High quality sequence stop: 293.
FEATURES             Location/Qualifiers
     source          1. .516
                     /organism="Mus musculus"
                     /note="Organ: bone marrow; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Not I - oligo(dT) primer
                     [5'
                     I and Eco RI sites of  the modified pT7T3 vector. Library
                     went through two rounds of  normalization, and was
                     constructed by Bento Soares and M.Fatima Bonaldo."
                     /db_xref="taxon:10090"
                     /clone="1749698"
                     /clone_lib="Soares mouse embryonic stem cell NMES"
                     /cell_type="embryonic stem cell"
                     /lab_host="DH10B"
BASE COUNT      157 a    141 c    130 g     85 t      3 others
ORIGIN
```

```
alignment_scores:
            Quality:   74.50              Length:     74
              Ratio:  1.863                Gaps:      5
 Percent Similarity:  54.054   Percent Identity: 29.730

alignment_block:
 US-08-487-032A-764 x AI048339/rev   ..

 Align seg 1/1  to reverse of: AI048339  from: 1  to: 516

      85 LeuAsnGlyPheGlyLeuAsnValGlyTyrLysLysPhePheGlnPheLy 101
         |||:::|||||||:::|||                  :::   |||::
     372 CTACACGGTTTTAGCCTT.....................CTTGGGTTTAG 344

     101 sSerLeuAspMetThrSerLysTrpPheGlyPheArgValTyrGlyLeuP 118
         ::::|||||||           |||||||:::|||         |||||||
     343 AGGCCTTGAC...........TGGTTTGACTTTGACAACCTTGGGCTTT 306

     118 heAspTyrGlyHisAlaAspLeuGlyLysGlnValTyrAlaProAsnLys 134
         ||::::::             |||||||:::                 ::::::
     305 TTGGCTTT...........CTTGGGCGT...........GGCAGCCGG 280

     135 IleGlnLeuAspMet.....................ValSerTrpGl 143
         :::   |||::::::                     ::::::|||||
     279 CTTCTTCTTGGCCTTCTTGACAGGGGTGGCTTTGGGTTTCTTGCTTGGGG 230

     143 yValGlySer***LeuLeuAla 150
         |   |||:::||||||||||:::
     229 CTTTGGAGGCAGCCTTCTTGGG 208
```

seq_name: gb_est15:AA713321

```
seq_documentation_block:
LOCUS        AA713321      469 bp    mRNA            EST       24-DEC-1997
DEFINITION   32881 Lambda-PRL2 Arabidopsis thaliana cDNA clone 249J4T7, mRNA
             sequence.
ACCESSION    AA713321
NID          g2723238
KEYWORDS     EST.
SOURCE       thale cress.
  ORGANISM   Arabidopsis thaliana
             Eukaryotae; Viridiplantae; Charophyta/Embryophyta group;
             Embryophyta; Tracheophyta; seed plants; Magnoliophyta;
             Magnoliopsida; Capparales; Brassicaceae; Arabidopsis.
REFERENCE    1  (bases 1 to 469)
  AUTHORS    Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H.,
             McIntosh,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M.,
             Retzel,E. and Somerville,C.
  TITLE      Genes galore: a summary of methods for accessing results from
             large-scale partial sequencing of anonymous Arabidopsis cDNA clones
  JOURNAL    Plant Physiol. 106, 1241-1255 (1994)
  MEDLINE    95148729
COMMENT
             Contact: Thomas Newman
             MSU-DOE Plant Research Laboratory
             Michigan State University
             MSU-DOE-PRL, Michigan State University,Plant Biology Bldg.,E.
             Lansing,Mi
             Tel: 517-353-0854
```

```
            Seq primer: T7.
 FEATURES             Location/Qualifiers
      source          1. .469
                      /organism="Arabidopsis thaliana"
                      /strain="var columbia"
                      /note="Vector: lambda Zip-Lox; Site_1: Sal; Site_2: Not;
                      Lambda PRL2 is a cDNA library derived from equal
                      quantities of 4 pools of mRNA.  The mRNA sources were 1) 7
                      day germinated etiolated seedlings; 2) tissue culture
                      grown roots; 3) staged plants half with 24 hour light
                      cycle, half on 16 hr light, 8 hour dark- rosettes;  4)
                      same plants as 3 but aerial tissue (stems, flowers and
                      siliques.  The vector is BRL's lambda Zip-Lox.  The cDNA
                      inserts were directionally cloned with Sal-Not arms using
                      oligo dT primed cDNA."
                      /db_xref="taxon:3702"
                      /clone="249J4T7"
                      /clone_lib="Lambda-PRL2"
 BASE COUNT      133 a      96 c     109 g     115 t      16 others
 ORIGIN


 alignment_scores:
             Quality:   73.00              Length:     26
               Ratio:  3.650                Gaps:      1
  Percent Similarity:  76.923   Percent Identity: 42.308

 alignment_block:
  US-08-487-032A-764 x AA713321/rev   ..

  Align seg 1/1  to reverse of: AA713321  from: 1  to: 469

      144 ValGlySer***LeuLeuAlaAspIleIle.Asp***ArgGlnArgPheP 160
          :::|||||||   |||   :::::::::||| :::   |||      ||||
      301 CTTGGGAGTCTTCTTGCCAGTAATGTTATCAGAGAGGAGGATGATCTTCT 252

      160 heTrpTyrPheTrpTrpGlyArgTyr 168
          |||||::::::||||||:::|||::: 
      251 TCTGGTGTCTGTGGTGGAACCGGTTT 226

 seq_name: gb_est13:C44138

 seq_documentation_block:
 LOCUS        C44138          372 bp     mRNA             EST       29-AUG-1997
 DEFINITION   C.elegans cDNA clone yk339e1 : 5' end, single read, mRNA sequence.
 ACCESSION    C44138
 NID          g2380375
 KEYWORDS     EST; EST(expressed sequence tag).
 SOURCE       Caenorhabditis elegans (strain:N2) embryo hermaphrodite embryo cDNA
              to mRNA, clone_lib:Yuji Kohara unpublished cDNA library
              clone:yk339e1.
   ORGANISM   Caenorhabditis elegans
              Eukaryotae; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
              Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 REFERENCE    1  (sites)
   AUTHORS    Kohara,Y., Motohashi,T., Tabara,H., Shin-i,T., Watanabe,H.,
              Sano,M., Miyata,A., Ohba,T., Mitani,Y., Uesugi,H., Sugiura,I.,
              Obara,M., Sugimoto,A., Iida,K. and Nishigaki,A.
   TITLE      Expression map of the C.elegans genome
   JOURNAL    Unpublished (1997)
 REFERENCE    2  (bases 1 to 372)
   AUTHORS    Kohara,Y.
   TITLE      Direct Submission
   JOURNAL    Submitted (23-JUL-1997) to the DDBJ/EMBL/GenBank databases. Yuji
              Kohara, National Institute of Genetics, Gene Network Lab; Yata
              1111, Mishima, Shizuoka 411, Japan (E-mail:ykohara@lab.nig.ac.jp,
              Tel:+81-559-81-6854, Fax:+81-559-81-6855)
 FEATURES             Location/Qualifiers
      source          1. .372
                      /organism="Caenorhabditis elegans"
                      /strain="N2"
                      /db_xref="taxon:6239"
                      /clone="yk339e1"
                      /clone_lib="Yuji Kohara unpublished cDNA library"
                      /dev_stage="embryo"
                      /sex="hermaphrodite"
                      /tissue_type="embryo"
 BASE COUNT      111 a      70 c      69 g     122 t
 ORIGIN


 alignment_scores:
             Quality:   72.00              Length:     13
               Ratio:  6.545                Gaps:      0
  Percent Similarity:  84.615   Percent Identity: 53.846

 alignment_block:
  US-08-487-032A-764 x C44138/rev   ..

  Align seg 1/1  to reverse of: C44138  from: 1  to: 372

      156 ArgGlnArgPhePheTrpTyrPheTrpTrpGlyArgTyr 168
          ||||||::::::||||||   :::||||||   |||:::
       51 CGACAAAAATATTTTTGGCCATATTGGTGGCCGAGGTTT 13

 seq_name: gb_est8:AA261060

 seq_documentation_block:
 LOCUS        AA261060        548 bp     mRNA             EST       18-MAR-1997
 DEFINITION   mz81d04.r1 Soares mouse NML Mus musculus cDNA clone 719815 5'
              similar to gb:X13171 Mouse mRNA for H1 histone subtype H1(0)
              (MOUSE);, mRNA sequence.
 ACCESSION    AA261060
 NID          g1897580
 KEYWORDS     EST.
 SOURCE       house mouse.
   ORGANISM   Mus musculus
              Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
              Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
              Mus.
 REFERENCE    1  (bases 1 to 548)
   AUTHORS    Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
              Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
              Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
              Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
              Waterston,R.
   TITLE      The WashU-HHMI Mouse EST Project
   JOURNAL    Unpublished (1996)
 COMMENT
              Contact: Marra M/Mouse EST Project
              WashU-HHMI Mouse EST Project
              Washington University School of MedicineP
              4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
              Tel: 314 286 1800
              Fax: 314 286 1810
              Email: mouseest@watson.wustl.edu
              This clone is available royalty-free through LLNL ; contact the
              IMAGE Consortium (info@image.llnl.gov) for further information.
              MGI:445311
              Seq primer: -28m13 rev2 ET from Amersham
              High quality sequence stop: 432.
 FEATURES             Location/Qualifiers
      source          1. .548
                      /organism="Mus musculus"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATCTTTTTTTTTTTTTTTTT 3'];
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      constructed and normalized by Bento Soares and M.Fatima
                      Bonaldo."
                      /db_xref="taxon:10090"
                      /clone="719815"
                      /clone_lib="Soares mouse NML"
```

```
                ASAAQLIGLIFWTLIWGRYLLALYLK"
     gene       1680. .4163
                /gene="HP0701"
     CDS        1680. .4163
                /gene="HP0701"
                /note="similar to EGAD:19444 percent identity: 97.45;
                identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="DNA gyrase, sub A (gyrA)"
                /db_xref="PID:g2313826"
                /translation="MQDNSVNETKNIVEVGIDSSIEESYLAYSMSVIIGRALPDARDG
                LKPVHRRILYAMHELGLTSKVAYKKSARIVGDVIGKYHPHGDNAVYDALVRMAQDFSM
                RLELVDGQGNFGSIDGDNAAAMRYTEARMTKASEEILRDIDKDTIDFVPNYDDTLKEP
                DILPSRLPNLLVNGANGIAVGMATSIPPHRMDEIIDALVHVLENPNAGLDEILEFVKG
                PDFPTGGIIYGKAGIIEAYKTGRGRVKVRAKVHVEKTKNKEIIVLDEMPFQTNKAKLV
                EQISDLAREKQIEGISEVRDESDREGIRVVIELKRDAMSEIVLNHLYKLTTMETTFSI
                ILLAIYNKEPKIFTLLELLHLFLNHRKTIIIRRTIFELEKAKARAHILEGYLIALDNI
                DEIVRLIKTSQSPEAAKNALMERFTLSEIQSKAILEMRLQRLTGLERDKIKEEYQNLL
                ELIDDLNGILKSEDRLNGVVKTELLEVKEQFSSPRRTEIQESYENIDIEDLIANEPMV
                VSMSYKGYVKRVDLKAYEKQNRGGKGKLSGSTYEDDFIENFFVANTHDILLFITNKGQ
                LYHLKVYKIPEASRIAMGKAIVNLISLAPDEKIMATLSTKDFSDERSLAFFTKNGVVK
                RTNLSEFESNRSCGIRAIVLDEGDELVSAKVVDKNAKHLLIASHLGIFIKFPLEEVRE
                IGRTTRGVIGIKLNENDFVVGAVVISDDGNKLLSVSENGLGKQTLAEAYRGQSRGGKG
                VIGMKLTQKTGNLVGVISVDDENLDLMILTASAKMIRVSIKDIRETGRNASGVKLINT
                ADKVMYVNSCPKEEEPENLETSSAQNLFE"
     gene       4160. .4636
                /gene="HP0702"
     CDS        4160. .4636
                /gene="HP0702"
                /note="hypothetical protein; identified by GeneMark;
                putative"
                /codon_start=1
                /transl_table=11
                /product="H. pylori predicted coding region HP0702"
                /db_xref="PID:g2313831"
                /translation="MMRFFSFFYFLFYFLGVSLQALSPLEDQEFLISYRLKIVDSRVM
                GEEYSVSKPIVSRIKTAPYVLDYHCSIITRNLPNLKNPLLPIKLERFLLEIALKKEKE
                RVIDCILKSQVAITHYDHSYKNGTTTTSILALKALSVRASLVGDALFLDIFRKEEE"
     gene       4633. .5778
                /gene="HP0703"
     CDS        4633. .5778
                /gene="HP0703"
                /note="similar to EGAD:38658 percent identity: 44.20;
                identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="response regulator"
                /db_xref="PID:g2313827"
                /translation="MKIAIVEDDINMRKSLELFFELQDDLEIVSFKNPKDALAKLDES
                FDLVITDINMPHMDGLEFLRLLEGKYESIVITGNATLNKAIDSIRLGVKDFFQKPFKP
                ELLLESIYRTKKVLEFQKKHPLEKPLKKPHKHSFLAASKALEESKRQALKVASTDANV
                MLLGESGVGKEVFAHFIHQHSQRSKHPFIAINMSAIPEHLLESELFGYQKGAFTDATA
                PKMGLFESANKGTIFLDEIAEMPLQLQSKLLRVVQEKEITRLGDNKSVKIDVRFISAT
                NANMKEKIAAKEFREDLFFRLQIVPITIAPLRERVEEILPIAEIKLKEVCDAYHLGPK
                SFSKNAAKCLLEYSWHGNVRELLGVVERAAILSEETEIQEKDLFLER"
     gene       6239. .6349
                /gene="HP0704"
     CDS        6239. .6349
                /gene="HP0704"
                /note="hypothetical protein; identified by GeneMark;
                putative"
                /codon_start=1
                /transl_table=11
                /product="H. pylori predicted coding region HP0704"
                /db_xref="PID:g2313833"
                /translation="MVSNSLNFFSNSHAIDFKGMASLLTLIYIINGYCCR"
     gene       complement(6450. .9257)
                /gene="HP0705"
     CDS        complement(6450. .9257)
                /gene="HP0705"
                /note="similar to EGAD:28894 percent identity: 53.44;
                identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="excinuclease ABC subunit A (uvrA)"
                /db_xref="PID:g2313828"
                /translation="MDKIIIQGARENNLKNIFLEIPKNQFVVFTGLSGSGKSTLAFDT

alignment_scores:
          Quality:   73.00              Length:     73
            Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-08-487-032A-764 x HPAE000583

 Align seg 1/1  to: HPAE000583  from: 1  to: 10378

       5 LysLysPheValAlaLeuGlyLeuLeuSerAlaValLeuSerSerSerLe 21
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    9452 AAAAAGTTTGTAGCTTTAGGGCTTCTATCCGCAGTTTTAAGCTCTTCGTT 9501

      21 uLeuAlaGluGlyAspGlyValTyrIleGlyThrAsnTyrGlnLeuGlyG 38
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    9502 GTTAGCCGAAGGTGATGGTGTTTATATAGGGACTAATTATCAGCTTGGAC 9551

      38 lnAlaArgLeuAsnSerAsnIleTyrAsnThrGlyAspCysThrGlySer 54
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    9552 AAGCCCGTTTGAATAGTAATATTTATAATACAGGGGATTGCACAGGGAGT 9601

      55 ValValGlyCysProProGlyLeuThrAlaAsnLysHisAsnProGlyGl 71
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    9602 GTTGTAGGTTGCCCCCCAGGTCTTACCGCTAATAAGCATAATCCAGGAGG 9651

      71 yThrAsnIleAsnTrpHis 77
         ||||||||||||||||||
    9652 CACCAATATCAATTGGCAT 9670

 seq_name: gb_pl1:ATU63815

 seq_documentation_block:
 LOCUS       ATU63815    63093 bp    DNA              PLN       12-SEP-1996
 DEFINITION  Arabidopsis thaliana AT.I.24-1, AT.I.24-2, AT.I.24-3, AT.I.24-4,
             AT.I.24-5, AT.I.24-6, AT.I.24-9 and AT.I.24-14 genes, partial cds,
             AT.I.24-7, ascorbate peroxidase (ATHAPX1), EF-1alpha-A1, -A2 and
             -A3 (EF-1alpha) and AT.I.24-13 genes, complete cds.
 ACCESSION   U63815
 NID         g1532162
 KEYWORDS    .
 SOURCE      thale cress.
   ORGANISM  Arabidopsis thaliana
             Eukaryotae; mitochondrial eukaryotes; Viridiplantae;
             Charophyta/Embryophyta group; Embryophyta; Magnoliophyta;
             Magnoliopsida; Capparales; Brassicaceae; Arabidopsis.
 REFERENCE   1  (bases 1 to 63093)
   AUTHORS   Axelos,M., Bardet,C., Liboz,T., Le Van Thai,A., Curie,C. and
             Lescure,B.
   TITLE     The gene family encoding the Arabidopsis thaliana translation
             elongation factor EF-1 alpha: molecular cloning, characterization
             and expression
   JOURNAL   Mol. Gen. Genet. 219 (1-2), 106-112 (1989)
   MEDLINE   90136496
 REFERENCE   2  (bases 1 to 63093)
   AUTHORS   Tremousaygue,D., Bardet,C., Dabos,P., Regad,F., Pelese,F. and
             Lescure,B.
   TITLE     Direct Submission
   JOURNAL   Submitted (15-JUL-1996) CNRS-INRA, LBMRPM, Chemin de Borde Rouge
             BP27, Castanet Tolosan 31326, France
 FEATURES             Location/Qualifiers
      source          1. .63093
                      /organism="Arabidopsis thaliana"
                      /strain="Ecotype Columbia"
                      /db_xref="taxon:3702"
                      /chromosome="I"
                      /map="position 24"
```

```
                     /gene="EF-1alpha"
     CDS             join(40037. .40498,40592. .41479)
                     /gene="EF-1alpha"
                     /note="translation elongation factor"
                     /codon_start=1
                     /product="EF-1alpha-A1"
                     /db_xref="PID:g1532172"
                     /translation="MGKEKFHINIVVIGHVDSGKSTTTGHLIYKLGGIDKRVIERFEK
                     EAAEMNKRSFKYAWVLDKLKAERERGITIDIALWKFETTKYYCTVIDAPGHRDFIKNM
                     ITGTSQADCAVLIIDSTTGGFEAGISKDGQTREHALLAFTLGVKQMICCCNKMDATTP
                     KYSKARYDEIIKEVSSYLKKVGYNPDKIPFVPISGFEGDNMIERSTNLDWYKGPTLLE
                     ALDQINEPKRPSDKPLRLPLQDVYKIGGIGTVPVGRVETGMIKPGMVVTFAPTGLTTE
                     VKSVEMHHESLLEALPGDNVGFNVKNVAVKDLKRGYVASNSKDDPAKGAANFTSQVII
                     MNHPGQIGNGYAPVLDCHTSHIAVKFSEILTKIDRRSGKEIEKEPKFLKNGDAGMVKM
                     TPTKPMVVETFSEYPPLGRFAVRDMRQTVAVGVIKSVDKKDPTGAKVTKAAVKKGAK"
     CDS             complement(join(44376. .45263,45359. .45820))
                     /gene="EF-1alpha"
                     /note="translation elongation factor"
                     /codon_start=1
                     /product="EF-1alpha-A2"

alignment_scores:
            Quality:    9.00            Length:      9
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-08-487-032A-764 x ATU63815/rev    ..

 Align seg 1/1  to reverse of: ATU63815  from: 1  to: 63093

      15 AlaValLeuSerSerSerLeuLeuAla 23
         ||||||||||||||||||||||||||||
   37413 GCTGTGCTAAGTTCGAGTTTGTTAGCT 37387

seq_name: gb_pr2:AC004656

seq_documentation_block:
LOCUS        AC004656    74155 bp    DNA             PRI       02-JUL-1998
DEFINITION   Homo sapiens Xp22 BAC GSHB-358P8 (Genome Systems Human BAC library)
             complete sequence.
ACCESSION    AC004656
NID          g3253114
KEYWORDS     HTG.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
             Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 74155)
  AUTHORS    Muzny,D., Arenson,A.D., Adams,C., Brundage,E., Bunac,C.,
             Carvelli,K., Chacko,J., Chen,J., Di,W., Ding,Y., Dugan,S.,
             Durbin,J., Forcum,J., Ganesh,R., Garcia,C., Goodman,M.,
             Gorrell,J.H., Haywood,M., Hernandez,J., Jackson,L., Jin,S.,
             Kampal,R., Karpathy,S., Kovar,C., Leal,B., Li,Y., Lichtarge,O.,
             Liu,W., Logan,O., Lu,J., Ly,T., Martinez,C., Oswal,G., Perez,L.,
             Rashid,N.D., Rowland,K., Savage,L., Scherer,S.E., Shen,H.,
             Simon,M., Stovall,K., Timms,K.M., Todd,J., Vo,Q., Williamson,A.,
             Worley,K.C., Yu,W., Chinault,C., Nelson,D. and Gibbs,R.A.
  TITLE      Direct Submission
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 74155)
  AUTHORS    Worley,K.C.
  TITLE      Direct Submission
  JOURNAL    Submitted (02-MAY-1998) Molecular and Human Genetics, Baylor
             College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE    3  (bases 1 to 74155)
  AUTHORS    Worley,K.C.
  TITLE      Direct Submission
  JOURNAL    Submitted (25-JUN-1998) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
REFERENCE    4  (bases 1 to 74155)
  AUTHORS    Worley,K.C.
  TITLE      Direct Submission
  JOURNAL    Submitted (02-JUL-1998) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
COMMENT      On Jun 25, 1998 this sequence version replaced gi:3242666.
             Sequencing is completed to a minimum standard of double strand
             coverage with a minimum of 2 clones and 2 reads with no ambiguities
             or 2 chemistries with a minimum of 2 clones and 3 reads with no
             ambiguities. If the sequence quality does not meet this standard,
             it will be indicated in the annotation.

             The repeat regions shown were identified using RepeatMasker by
             Adrian Smit.

             Sequence similarities were identified using Powerblast by Jinghui
             Zhang.

             Exon/Intron boundaries of identified genes were chosen if there
             were canonical splice junctions that maintained sequence continuity
             across the splice junctions.
FEATURES             Location/Qualifiers
     source          1. .74155
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="GSHB-358P8"
                     /chromosome="X"
                     /clone_lib="Genome Systems Human BAC library"
                     /map="Xp22"
     misc_feature    1. .206
                     /note="overlap bases 137066 to 137270 in AC004552."
     repeat_region   complement(282. .561)
                     /rpt_family="AluSx"
     repeat_region   878. .1054
                     /rpt_family="MIR"
     repeat_region   complement(1783. .2080)
                     /rpt_family="AluJb"
     repeat_region   complement(2087. .2380)
                     /rpt_family="AluSx"
     repeat_region   complement(2384. .2425)
                     /rpt_family="L2"
     repeat_region   complement(2924. .3108)
                     /rpt_family="L1MC4"
     repeat_region   3138. .3439
                     /rpt_family="AluSx"
     misc_feature    3808. .4265
                     /note="Region: Similar to N48206 yv22b03.r1 Soares fetal
                     liver spleen 1NFLS Homo sapiens cDNA clone 243437 5'"
     repeat_region   4150. .4291
                     /rpt_family="MLT1G"
     repeat_region   complement(4703. .4986)
                     /rpt_family="AluJo"
     repeat_region   5404. .5576
                     /rpt_family="L1PA16"
     misc_feature    complement(5425. .5939)
                     /note="Region: N39012 yv22b03.s1 Soares fetal liver spleen
                     1NFLS Homo sapiens cDNA clone 243437 3'"
     repeat_region   5987. .6096
                     /rpt_family="AluSp/q"
     repeat_region   6479. .6785
                     /rpt_family="AluSx"
     repeat_region   7475. .7776
                     /rpt_family="AluSx"
     repeat_region   complement(8221. .8347)
                     /rpt_family="FLAM_A"
     repeat_region   complement(8989. .9062)
                     /rpt_family="L1MC4"
     repeat_region   9113. .9370
                     /rpt_family="AluSx"
     repeat_region   complement(9374. .9745)
                     /rpt_family="MER31A"
     repeat_region   9746. .10021
                     /rpt_family="AluJb"
     repeat_region   complement(10022. .10069)
```

```
seq_name: N_Geneseq_34:Q12746

seq_documentation_block:
ID   Q12746 standard; DNA; 639 BP.
AC   Q12746;
DT   27-SEP-1991  (first entry)
DE   B.burgdorferi strain PKo pC gene.
KW   lyme borreliosis; vaccine; flagellin; ss.
OS   Borrelia burgdorferi.
PN   WO9109870-A.
PD   11-JUL-1991.
PF   21-DEC-1990; E02282.
PR   22-DEC-1989; DE-942728.
PR   13-JUN-1990; DE-018988.
PA   (MIKR-) MIKROGEN MOLEKULARB.
PI   Fuchs R, Wilske B, Preac-Mursic V, Motz M, Soutschek E;
DR   WPI; 91-222844/30.
DR   P-PSDB; R13140.
PT   New Borrelia burgdorferi proteins - useful as immunoassay
PT   reagents and antigens for vaccine prodn.
PS   Example 3; Page 26; 68pp; German.
CC   Protein pC was isolated from a B.burgdorferi cell lysate and
CC   digested with trypsin. The amino acid sequence of two tryptic
CC   fragments was determined. Probe pools corresponding to each
CC   fragment were synthesised and used to screen a B.burgdorferi cDNA
CC   library. A clone contg. the 639 nucleotides of the pC coding
CC   sequence was identified and sequenced. The protein sequence
CC   decoded from this coding sequence does not correspond to the amino
CC   acid sequence printed in the specification (R13140). For the two
CC   sequences to correspond, insert an A residue between G(84) and
CC   C(85) and delete T(111).
CC   See Q12744-Q12747, Q13297-8 and R13139-R13142.
SQ   Sequence   639 BP;    252 A;    88 C;    116 G;    183 T;


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-08-487-032A-764 x Q12746   ..

 Align seg 1/1  to: Q12746  from: 1  to: 639

     121 GlyHisAlaAspLeuGlyLysGln 128
         ||||||||||||||||||||||||||
     427 GGTCATGCAGATCTTGGCAAACAG 450

seq_name: N_Geneseq_34:Q65405

seq_documentation_block:
ID   Q65405 standard; cDNA; 1670 BP.
AC   Q65405;
DT   30-NOV-1994  (first entry)
DE   Cuphea hookeriana thioesterase cDNA clone CMT10.
KW   Cuphea hookeriana; tree; ACP-thioesterase;
KW   acyl carrier protein-thioesterase; enzyme; ds.
OS   Cuphea hookeriana.
PN   WO9410288-A.
PD   11-MAY-1994.
PF   29-OCT-1993; U10814.
PR   30-OCT-1992; US-968971.
PA   (CALJ ) CALGENE INC.
PI   Davies HM,  Knutzon DS,  Voelker TA;
DR   WPI; 94-167447/20.
PT   DNA constructs encoding medium- and long-chain acyl-ACP
PT   thioesterases - useful commercially as detergents, lubricants,
PT   and in sports and low calorie foods
PS   Disclosure; Fig.9; 79pp; English.
CC   This DNA encodes a Cuphea hookeriana thioesterase cDNA clone CMT10.
CC   activity towards fatty acyl-ACPs.
SQ   Sequence   1670 BP;    408 A;    375 C;    436 G;    451 T;


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-08-487-032A-764 x Q65405/rev   ..

 Align seg 1/1  to reverse of: Q65405  from: 1  to: 1670

      14 SerAlaValLeuSerSerSerLeu 21
         |||||||||||||||||||||||
    1115 TCAGCAGTCTTGTCATCCAGCTTG 1092

seq_name: N_Geneseq_34:Q73869

seq_documentation_block:
ID   Q73869 standard; DNA; 585 BP.
AC   Q73869;
DT   25-MAY-1995  (first entry)
DE   Borrelia JSB antigen vaccine.
KW   OspC antigen; vaccine; Lyme disease; borreliosis; immunogen;
KW   serovar typing; restriction fragment length polymorphism;
KW   RFLP; Pichia pastoris; ss.
OS   Borrelia burgdorferi JSB.
PN   WO9425596-A.
PD   10-NOV-1994.
PF   29-APR-1994; E01365.
PR   29-APR-1993; US-053863.
PA   (IMMO ) IMMUNO AG.
PI   Crowe B,  Dorner F,  Livey I;
DR   WPI; 94-358273/44.
DR   P-PSDB; R60896.
PT   Immunogenic composition comprising OspC antigens - for the
PT   treatment of Lyme borreliosis in different, specific geographical
PT   areas.
PS   Disclosure; Fig. 8a; 115pp; English.
CC   A vaccine for Lyme disease includes selected OspC antigen
CC   formulations based on defined OspC families resolved by serovar
CC   typing and RFLP typing.  Partial sequences of OspC genes selected
CC   from different RFLP types are given in Q73883-905 (encoded peptides,
CC   comprising the first 92% of mature OspC, are given in R62771-93).
CC   Complete sequences of these novel ospC genes, including the 3' end,
CC   plus sequences for the ospC genes of Borrelia strains H13 and 28691
CC   are given in Q73857-82, and encoded proteins in R60884-909.  The
CC   DNA sequences may be expressed in e.g. Pichia pastoris for
CC   recombinant antigen production.
SQ   Sequence   585 BP;    231 A;    85 C;    110 G;    159 T;


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-08-487-032A-764 x Q73869   ..

 Align seg 1/1  to: Q73869  from: 1  to: 585

     121 GlyHisAlaAspLeuGlyLysGln 128
         |||||||||||||||||||||||||
     373 GGTCATGCAGATCTTGGCAAACAG 396

seq_name: N_Geneseq_34:Q73894

seq_documentation_block:
ID   Q73894 standard; DNA; 531 BP.
```

```
DT   25-MAY-1995  (first entry)
DE   Borrelia JSB antigen vaccine.
KW   OspC antigen; vaccine; Lyme disease; borreliosis; immunogen;
KW   serovar typing; restriction fragment length polymorphism;
KW   RFLP; Pichia pastoris; ss.
OS   Borrelia burgdorferi JSB.
PN   WO9425596-A.
PD   10-NOV-1994.
PF   29-APR-1994; E01365.
PR   29-APR-1993; US-053863.
PA   (IMMO ) IMMUNO AG.
PI   Crowe B,  Dorner F,  Livey I;
DR   WPI; 94-358273/44.
DR   P-PSDB; R62782.
PT   Immunogenic composition comprising OspC antigens - for the
PT   treatment of Lyme borreliosis in different, specific geographical
PT   areas.
PS   Disclosure; Fig. 8; 115pp; English.
CC   A vaccine for Lyme disease includes selected OspC antigen
CC   formulations based on defined OspC families resolved by serovar
CC   typing and RFLP typing.  Partial sequences of OspC genes selected
CC   from different RFLP types are given in Q73883-905 (encoded peptides,
CC   comprising the first 92% of mature OspC, are given in R62771-93).
CC   Complete sequences of these novel ospC genes, including the 3' end,
CC   plus sequences for the ospC genes of Borrelia strains H13 and 28691
CC   are given in Q73857-82, and encoded proteins in R60884-909.  The
CC   DNA sequences may be expressed in e.g. Pichia pastoris for
CC   recombinant antigen production.
SQ   Sequence   531 BP;    208 A;    76 C;    100 G;    147 T;


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-08-487-032A-764 x Q73894    ..

 Align seg 1/1  to: Q73894  from: 1  to: 531

     121 GlyHisAlaAspLeuGlyLysGln 128
         ||||||||||||||||||||||||||
     370 GGTCATGCAGATCTTGGCAAACAG 393

 seq_name: N_Geneseq_34:T17844

 seq_documentation_block:
 ID   T17844 standard; DNA; 1688 BP.
 AC   T17844;
 DT   24-MAY-1996  (first entry)
 DE   Soybean palmitoyl-ACP-thioesterase gene clone pTE11.
 KW   Soybean; palmitoyl-ACP-thioesterase; fatty acid; lipid;
 KW   vegetable oil; palmitic acid; stearic acid; triacylglycerol;
 KW   transgenic plant; ss.
 OS   Glycine max.
 FH   Key             Location/Qualifiers
 FT   transit_peptide 488. .505
 FT                   /*tag=  a
 FT   mat_peptide     506. .1477
 FT                   /*tag=  b
 FT   primer_bind     complement (1. .16)
 FT                   /*tag=  c
 FT                   /note= "primer SOYTE3"
 FT   primer_bind     1640. .1657
 FT                   /*tag=  d
 FT                   /note= "primer SOYTE4"
 FT
 PN   WO9606936-A1.
 PD   07-MAR-1996.
 PF   25-AUG-1995; U10627.
 PR   31-AUG-1994; US-299044.
 PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
 PI   Hitz WD;
 DR   WPI; 96-160367/16.
 PT   Canola and soybean palmitoyl-ACP-thio:esterase genes - useful in
 PT   regulation of fatty acid content of the oils of canola and soybean
 PT   plants
 PS   Claim 2; Page 69-70; 103pp; English.
 CC   A cDNA clone (T17845), designated pTE11, codes for soybean
 CC   palmitoyl-ACP-thioesterase.  It was obtd. from a Glycine max
 CC   embryo cDNA library by screening with a probe based on an Arabidopsis
 CC   thioesterase-like fragment.  A restriction fragment of pTE11 was
 CC   expressed in Escherichia coli BL21(DE3)(pLysE) cells.  The construct
 CC   encoded a 328-amino acid protein (R82789).  Chimeric genes including
 CC   the isolated sequence can be used to create transgenic plants having
 CC   altered levels of saturated fatty acids.  Vegetable oils (esp.
 CC   soybean or rapeseed) may be produced that contain lower or higher
 CC   than normal levels of stearic acid and palmitic acid.
 SQ   Sequence   1688 BP;    459 A;    349 C;    395 G;    485 T;


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-08-487-032A-764 x T17844/rev    ..

 Align seg 1/1  to reverse of: T17844  from: 1  to: 1688

      14 SerAlaValLeuSerSerSerLeu 21
         |||||||||||||||||||||||||
    1119 TCCGCTGTGTTGTCGTCAAGTTTA 1096

 seq_name: N_Geneseq_34:Q90716

 seq_documentation_block:
 ID   Q90716 standard; DNA; 639 BP.
 AC   Q90716;
 DT   31-JUL-1996  (first entry)
 DE   B. burgdorferi strain PKo outer surface protein C (OspC-PKo) DNA.
 KW   Strain PKo; outer surface protein; OspC; antigenic domain;
 KW   chimaeric protein; treatment; diagnosis; infection; vaccine;
 KW   Lyme borreliosis; immunodiagnostic assay; antibody;
 KW   T-cell reactivity; chimeric; ds.
 OS   Borrelia burgdorferi.
 FH   Key             Location/Qualifiers
 FT   cds             1. .639
 FT                   /*tag=  a
 FT
 PN   WO9512676-A1.
 PD   11-MAY-1995.
 PF   27-OCT-1994; U12352.
 PR   01-NOV-1993; US-148191.
 PR   29-APR-1994; US-235836.
 PA   (ASUY-) ASSOC UNIVERSITIES INC.
 PI   Dunn JJ,  Luft BJ;
 DR   WPI; 95-215034/28.
 DR   P-PSDB; R75729.
 PT   Chimeric protein comprising 2 or more antigenic Borrelia
 PT   polypeptide(s) - useful in a vaccine against Lyme borreliosis and in
 PT   immuno:diagnostic assays
 PS   Example 1; Fig 14; 200pp; English.
 CC   The present sequence encodes the B. burgdorferi strain PKo, outer
 CC   surface protein C (OspC-PKo). Using chemical or enzymatic methods,
 CC   peptide fragments of OspC-PKo were prepd., and analysed by western
 CC   blot to assess their ability to bind different anti-OspC monoclonal
 CC   antibodies. The information obtd. was used to locate antigenic
 CC   domains in OspC-PKo, the epitopes of which were mapped with the
 CC   aid of site directed mutagenesis. Identical analyses were performed
 CC   on a selection of Osp purified from a variety of B. burgdorferi
 CC   strains, the results from which were utilised in the prepn. of a
 CC   pool of antigenic Borrelia polypeptides, and corresponding
 CC   polynucleotides. Chimaeric proteins comprising 2 or more antigenic
 CC   Borrelia polypeptides, that do not naturally occur in the same
 CC   protein, can be used in the treatment and diagnosis of Borrelia
```

```
 US-08-487-032A-764 x V65261/rev   ..

 Align seg 1/1  to reverse of: V65261  from: 1  to: 2766

     163 PheTrpTrpGlyArgTyrArgArg 170
         ||||||||||||||||||||||||
    2703 TTCTGGTGGGGAAGATATCGCAGA 2680

seq_name: N_Geneseq_34:N91093

seq_documentation_block:
ID   N91093 standard; DNA; 1950 BP.
AC   N91093;
DT   04-JUL-1990  (first entry)
DE   Protein G gene.
KW   Protein G; immunoglobulin; Fc receptor; ds.
OS   Streptococcus sp.
FH   Key             Location/Qualifiers
FT   -35_signal      465. .470
FT                   /*tag= c
FT   -10_signal      487. .492
FT                   /*tag= b
FT   rbs             565. .571
FT                   /*tag= d
FT   cds             578. .1918
FT                   /*tag= a
PN   WO8810306-A.
PD   29-DEC-1988.
PF   20-JUN-1988; 02084.
PR   19-JUN-1987; US-063959.
PA   (GENE-) Genex Corp.
PI   Fahnestock SR;
DR   WPI; 89-023848/03.
DR   P-PSDB; P95030.
PT   Cloned protein G variant genes -
PT   expressing proteins having immunoglobulin-binding properties of
PT   protein G and derived from Streptococcus sp.
PS   Disclosure; pp; English.
CC   Gene encodes protein G of non-pathogenic streptococcus sp. allowing
CC   isolation of the protein and variants, useful as bacterial Fc receptors
CC   eg in purification and detection of Abs., screening of hybridoma clones
CC   and treatment of disease.
SQ   Sequence   1950 BP;   705 A;   323 C;   398 G;   524 T;


alignment_scores:
            Quality:    7.00            Length:      7
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-08-487-032A-764 x N91093/rev   ..

 Align seg 1/1  to reverse of: N91093  from: 1  to: 1950

      17 LeuSerSerSerLeuLeuAla 23
         |||||||||||||||||||||
    1152 TTGTCAAGTTCTCTGTTAGCT 1132
```